Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1.   (Currently amended) A method performed by a computer system, the method comprising:

extracting, by a processor of the computer system, a set of uniform resource locators (URLs) from one document or from multiple documents associated with a single web host;

identifying, by the processor, sub-strings occurring in multiple URLs in the set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs;

generating, by the processor, a clean set of URLs from the set of URLs by removing the session identifiers; and

determining, by the processor, when at least one particular URL has already been crawled based on a comparison of the particular URL to the clean set of URLs.

2.   (Canceled)

3.   (Previously presented) The method of claim 1, where the document or each of the multiple documents is a web document downloaded from a web site.

4.   (Previously presented) The method of claim 1, where the comparison of the

particular URL to the clean set of URLs comprises calculating a fingerprint value for a particular

URL and for each of the URLs in the clean set of URLs, and where the comparison

is based on a comparison of the fingerprint value of the particular URL to the fingerprint values

of the URLs in the clean set of URLs.

5.   (Previously presented) The method of claim 1, where the particular rule

comprises:

     determining that the sub-strings do not reference content.

6.   (Canceled)

7.   (Previously presented) The method of claim 1, where the particular rule

comprises:

     determining that the sub-strings contain characters consistent with a session identifier.

8.   (Previously presented) The method of claim 1, further comprising:

     downloading content from the particular URL when the particular URL is determined to

not already have been crawled.

9.   (Previously presented) The method of claim 1, further comprising:

     storing information based on the clean set of URLs for use in later determining whether

additional URLs have already been extracted; and

storing the set of URLs, including embedded session identifiers, for use in later accessing the set of URLs.

10.   (Currently amended) A method performed by a computer system, the method comprising:

receiving, by a communication interface or an input device of the computer system, a set of uniform resource locators (URLs);

analyzing, by a processor of the computer system, the set of URLs for sub-strings that are structured in a manner consistent with session identifiers; and

further analyzing, by the processor, the set of URLs to identify one of the sub-strings as corresponding to a session identifier based on multiple occurrences of the sub-string in the set of URLs.

11.   (Previously presented) The method of claim 10, where the set of URLs are extracted from a web document associated with a web host.

12.   (Previously presented) The method of claim 10, where the set of URLs are extracted from multiple web documents associated with a single web host.

13.   (Previously presented) The method of claim 10, further comprising:

removing identified session identifiers from the set of URLs; and

storing the set of URLs, with the removed session identifiers, as a clean set of URLs.

14.   (Previously presented) The method of claim 13, further comprising:

adding a generated session identifier to URLs in the clean set of URLs.

15.   (Currently amended) A device comprising:

a memory to store instructions; and

a processor to execute the instructions to implement:

        at least one fetch bot ~~configured~~ to download content on a network from locations

specified by uniform resource locators (URLs);

        a content manager ~~configured~~ to:

                extract URLs from the downloaded content, and

                identify session identifiers from the URLs extracted from the downloaded

        content based, at least in part, on multiple occurrences of the session identifiers

        from a single web site; and

        a URL manager ~~configured~~ to create clean versions of the URLs extracted from

the downloaded content by removing the session identifiers from the URLs and to store

the clean versions of the URLs.

16.   (Currently amended) The device of claim 15, where the content manager is

further ~~configured~~ to identify the session identifiers based on locating sub-strings, within the

URLs extracted from the downloaded content, that contain characters consistent with session

identifiers.

17.   (Currently amended) The device of claim 15, further comprising:

a database ~~configured~~ to store the downloaded content.

18.   (Currently amended) The device of claim 15, where the URL manager is further ~~configured~~ to determine when additional URLs have previously been stored by comparing clean versions of the additional URLs to the stored clean versions of the URLs extracted from the downloaded content.

19.   (Previously presented) The device of claim 15, where the session identifiers include characters from the URLs extracted from the downloaded content that do not reference content.

20.   (Currently amended) A device comprising:

    hardware-implemented means for receiving a set of uniform resource locators (URLs);

    hardware-implemented means for analyzing the set of URLs for sub-strings that are structured in a manner consistent with session identifiers; and

    hardware-implemented means for further analyzing the set of URLs to identify one of the sub-strings as corresponding to a session identifier based on multiple occurrences of the sub-string in the set of URLs.

21.   (Previously presented) The device of claim 20, where the set of URLs are extracted from a web document associated with a web host.

22.   (Previously presented) The device of claim 20, where the set of URLs are extracted from multiple web documents associated with a single web host.

23.   (Original) The device of claim 20, further comprising:

means for removing the identified session identifiers from the set of URLs; and

means for storing the set of URLs with the removed session identifiers as a clean set of URLs.

24.   (Previously presented) The device of claim 23, further comprising:

means for adding a generated session identifier to URLs in the clean set of URLs.

25.   (Currently amended) ~~A computer-readable memory device including~~ One or more memory devices that include programming instructions that when executed by at least one processor causes the at least one processor to perform a method including:

receiving a set of uniform resource locators (URLs);

analyzing the set of URLs for sub-strings that are structured in a manner consistent with session identifiers; and

further analyzing the set of URLs to identify one of the sub-strings as corresponding to a session identifier based on multiple occurrences of the sub-string in the set of URLs.

26.   (Currently amended) The ~~computer-readable memory device~~ one or more memory devices of claim 25, where the set of URLs are extracted from a web document associated with a web host.

27.   (Currently amended) The ~~computer-readable memory device~~ one or more memory devices of claim 25, where the set of URLs are extracted from multiple web documents associated with a single web host.

28.   (Currently amended) The ~~computer-readable memory device~~ one or more memory devices of claim 25, where the programming instructions further include programming instructions that cause the at least one processor to:

    remove the session identifiers from the set of URLs; and

    store the set of URLs with the removed session identifiers as a clean set of URLs.

29.   (Currently amended) The ~~computer-readable memory device~~ one or more memory devices of claim 28, where the programming instructions further include programming instructions that cause the at least one processor to:

    add a generated session identifier to URLs in the clean set of URLs when the URLs are to be used to access a web document.